Harvesting Event Chains in Ritual Descriptions using Frame Semantics: Chances and Challenges

Anette Frank Nils Reiter

Department of Computational Linguistics
Heidelberg University, Germany

1st AMICUS Workshop, Vienna, September 30, 2010

Overview

Rituals

Offer a lamp and speak the mantra tejo 'si.

Shower pieces of fruits from the measuring vessel on the head of the boy with yāḥ phalini.

Investigating the Structure of Rituals

- Many rituals contain re-occurring building blocks
- Ritual elements are used to compose rituals
- The composition follows rules
- Is there a "Ritual Grammar"? Michaels (2007)
  - Can we identify the "ritual elements"?
  - Can we identify rules of composition?
  - Are there culture-independent rules or elements?

Project Aims

- Investigating the *event structure* of rituals
- Detecting *regularities, similarities and differences*
- Using *empirical, computational linguistics methods*

What does CL have to offer?

Event Annotation using Frame Semantics

- Frames: Prototypical events (e.g., TAKING)
- Frame Elements: Frame-specific semantic roles (e.g., AGENT, THEME, SOURCE)
- Frame Relations: Form a network of frames through inheritance, precedence, etc.
- Scenario Frames: Sequences of events forming a "scenario"

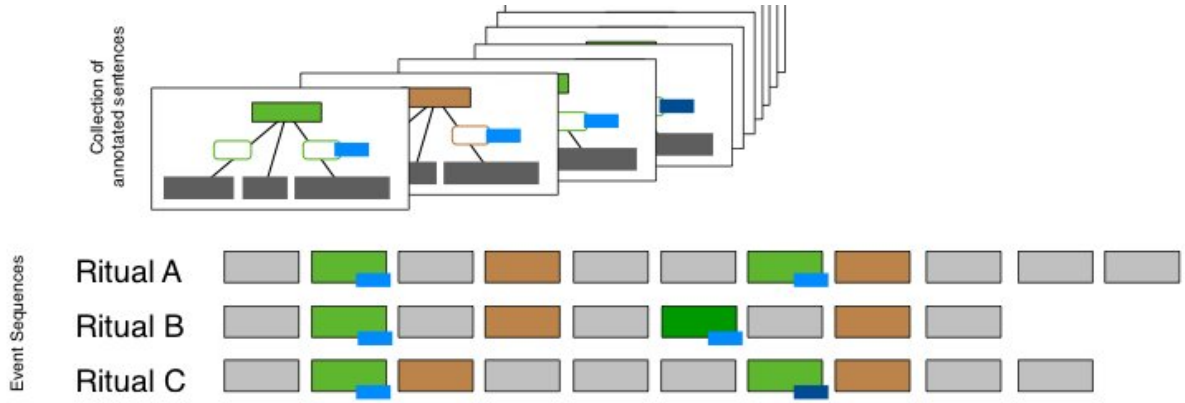
Sense Annotation of Frame Elements

- *Offer a lamp*.
  OIL LAMP > LAMP > DEVICE .. > OBJECT
- Computing selectional preferences and divergences

Harvesting Event Chains from Ritual Descriptions

Frame semantics as an abstraction layer for event chains

Burchardt et al. (2005)



Detecting structural elements

- patterns of (partial) event sequences Hellwig (2009)
- similarities and differences across event role fillers

Using sequence alignment and semantic distance measures

Overview

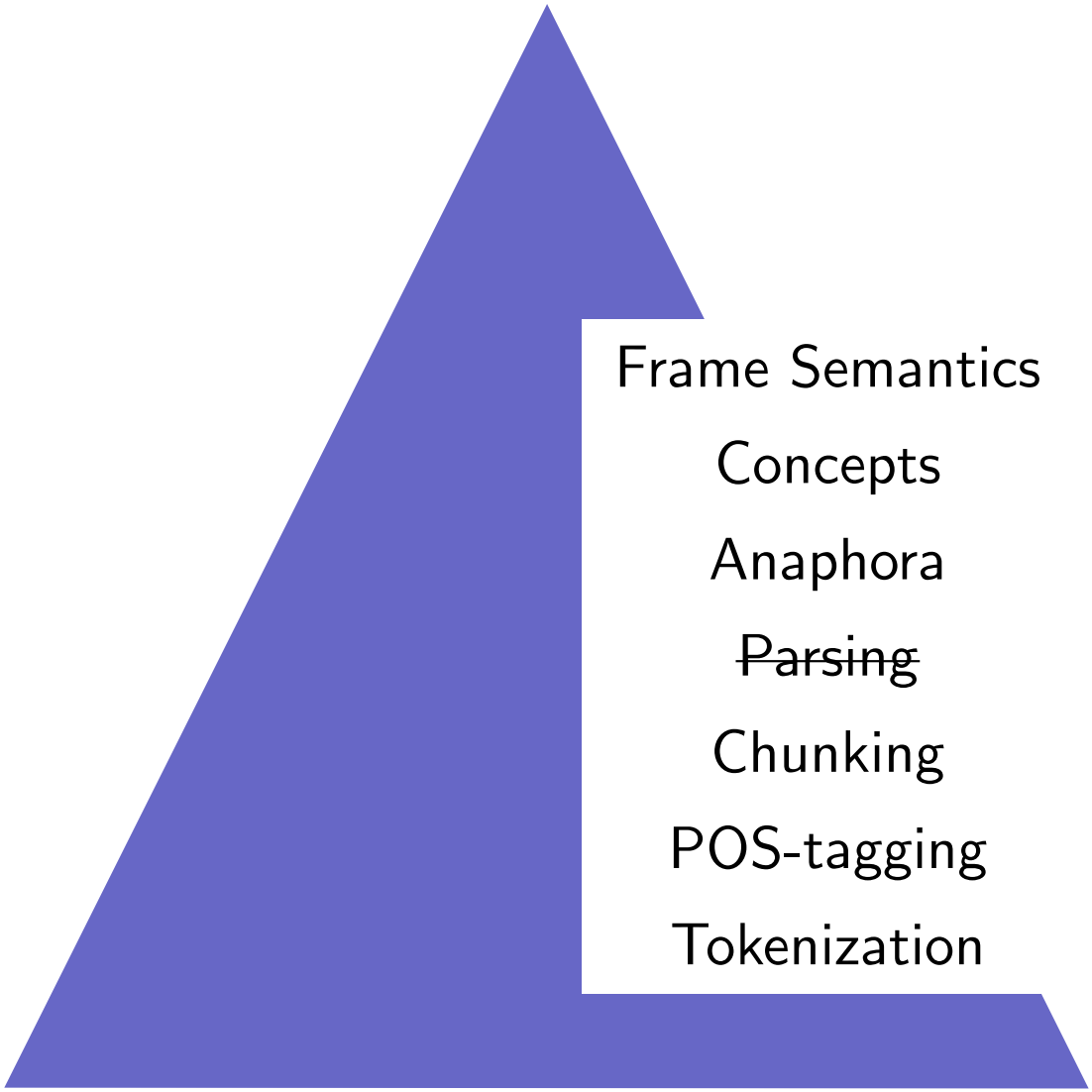
Data Sources

	Handbook (prescriptive) (practitioners)	Observation (descriptive) (researchers)
Initiation	12	8
Death and Ancestor	1	19

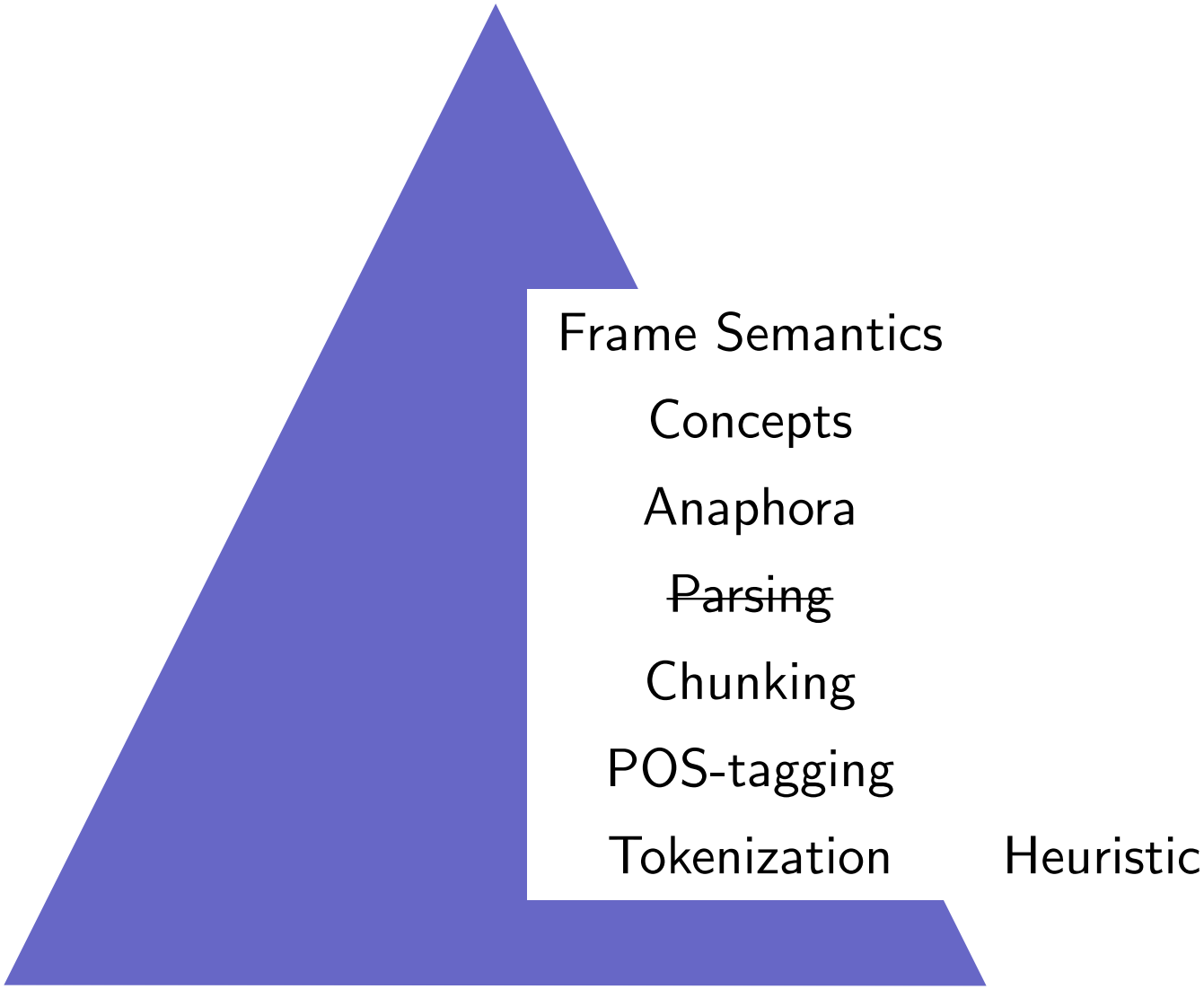
Table: Collected ritual descriptions

Michaels and Gutschow (2005; 2008)

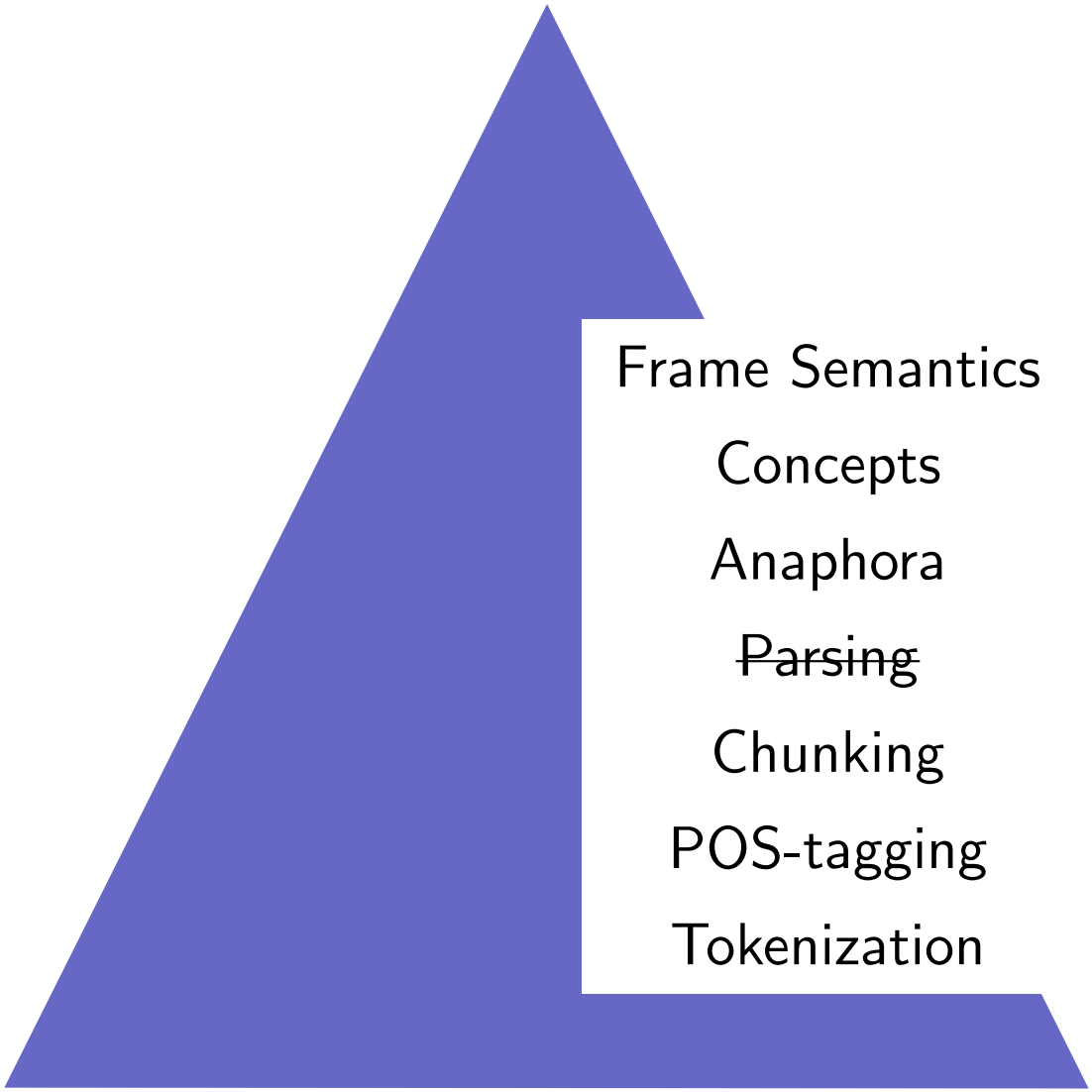
Annotation Pipeline



Annotation Pipeline



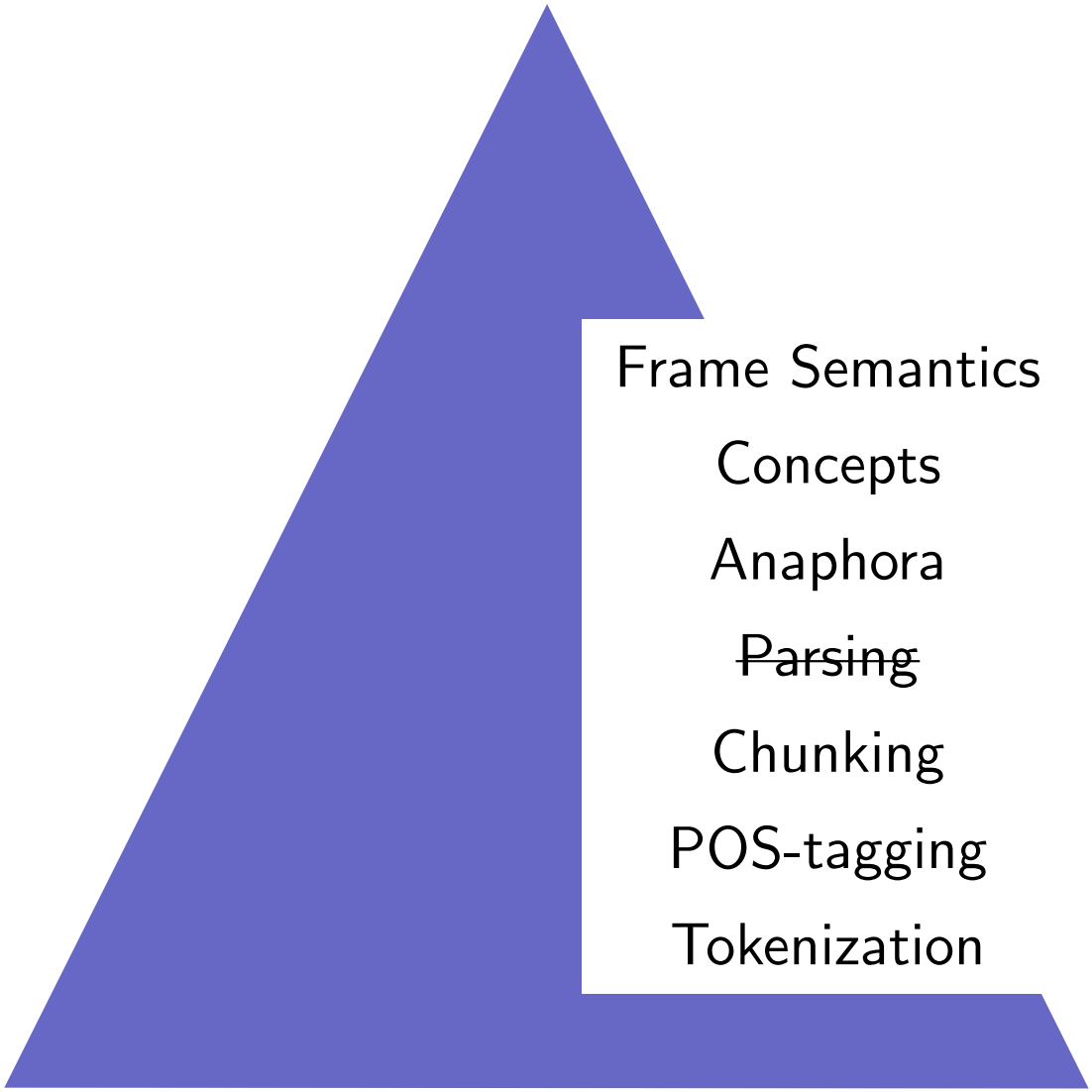
Annotation Pipeline



OpenNLP

Morton et al. (2005)

Annotation Pipeline



Manual Annotation

Overview

Text Characteristics

Ritual Descriptions have very special characteristics

- Foreign terms
- Fixed expressions
- Imperatives
- Complex sentence structure
- Interpretations

Reiter et al. (2010)

Text Characteristics

Ritual Descriptions have very special characteristics

- Foreign terms
- Fixed expressions
- Imperatives
- Complex sentence structure
- Interpretations

Example

Hand over the worship materials.

Reiter et al. (2010)

Text Characteristics

Ritual Descriptions have very special characteristics

- Foreign terms
- Fixed expressions
- Imperatives
- Complex sentence structure
- Interpretations

Example

Hand over the worship materials.

Frequency of imperative verbs

(BNC) 2% $\leftrightarrow$ 70% (Ritual texts)

Reiter et al. (2010)

Domain Adaptation

In general: Transfer models trained on one domain onto another

Standard Case

Source Domain Standard domain like newspaper texts;
Lots of training data available

Target Domain New domain like ritual descriptions;
No training data available

Annotating a *small* amount of target domain data is often feasible

Daumé III et al. (2010), Daumé III and Marcu (2006)

Domain Adaptation: Combining Data Sets

- Retraining of model(s) on various data sets

	Data	Description
baseline	WSJ	The Wall Street Journal
	Rit	Ritual descriptions
data set	WSJ + Rit	Union of WSJ and Rit
	WSJ + Rit $\uparrow$	Oversampling Rit
	WSJ $\downarrow$ + Rit	Undersampling WSJ
features	WSJ $\times$ Rit	Augmented features (Daumé III 2007)
	WSJ $\times$ Rit $\uparrow$	Oversampling Rit
	WSJ $\downarrow \times$ Rit	Undersampling WSJ

Table: Overview of domain adaptation approaches

Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the "general domain"

Daumé III (2007)

Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the "general domain"

Daumé III (2007)

Example

```
DT   The
NN   monitor
VB   is
VBG  flickering
```

```
PRP  We
VB   monitor
JJ   daily
NNS  news
```

Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the "general domain"

Daumé III (2007)

Example

DT	The
NN	monitor
VB	is
VBG	flickering

PRP	We
VB	monitor
JJ	daily
NNS	news

DT	The
NN	monitor
VB	is
VBG	flickering
PRP	We
VB	monitor
JJ	daily
NNS	news

The
monitor
is
flickering
We
monitor
daily
news

Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the "general domain"

Daumé III (2007)

Example

DT	The
NN	monitor
VB	is
VBG	flickering

PRP	We
VB	monitor
JJ	daily
NNS	news

DT	The	0The
NN	monitor	0monitor
VB	is	0is
VBG	flickering	0flickering
PRP	We	1We
VB	monitor	1monitor
JJ	daily	1daily
NNS	news	1news

Results for POS-Tagging

	Training data	Accuracy
	WSJ	94.01 %
	Rit	95.47 %
data set	WSJ + Rit	97.32 %
	WSJ + Rit $\uparrow$	97.59 %
	WSJ $\downarrow$ + Rit	96.97 %
features	WSJ $\times$ Rit	97.19 %
	WSJ $\times$ Rit $\uparrow$	97.40 %

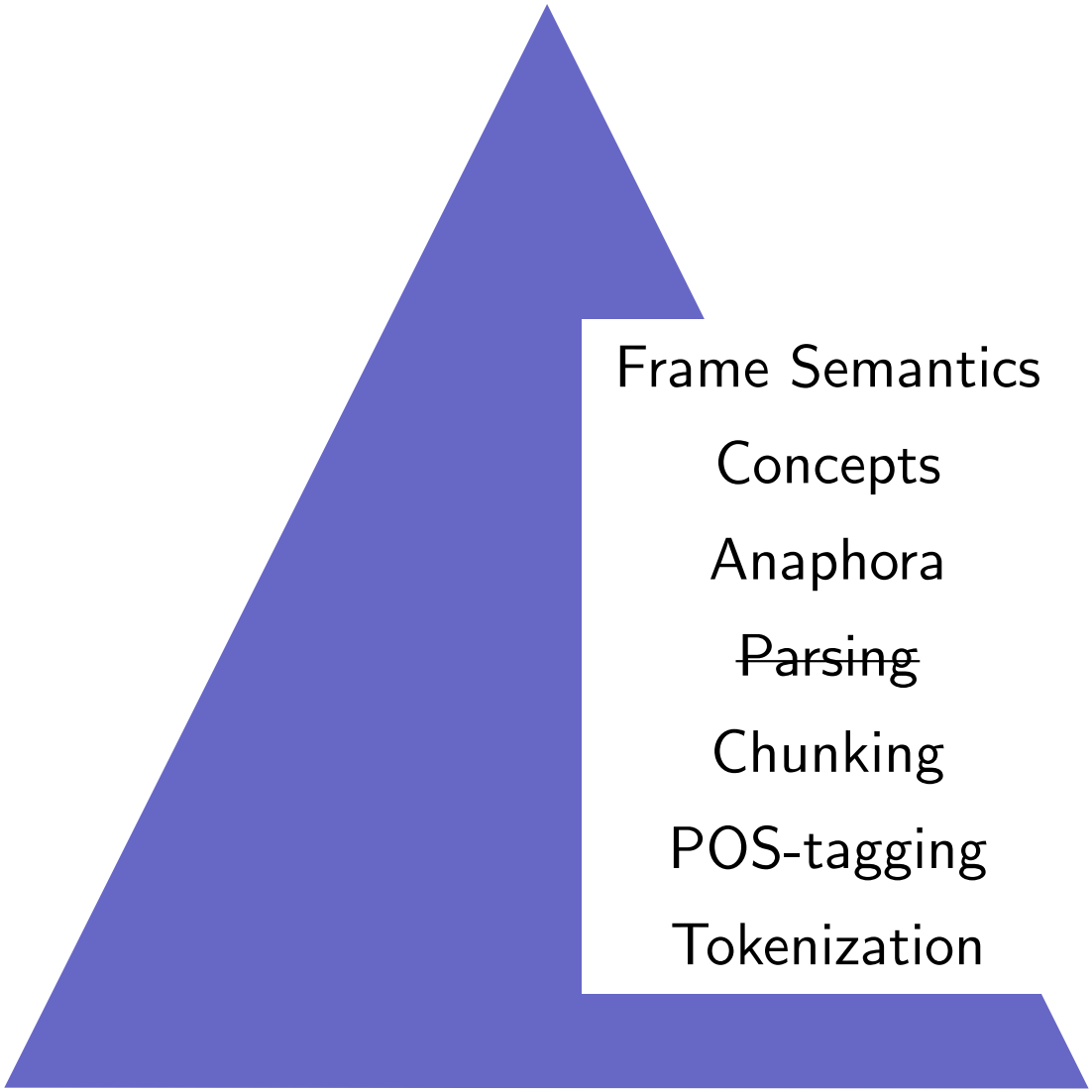
Table: Part of speech tagging results with different models. (Tested on 408 sentences using 10-fold cross validation)

Results for Chunking

	Training data	Precision	Recall	$F_{\beta=1}$
	WSJ	87.72 %	87.23 %	87.47
	Rit	91.09 %	89.85 %	90.47
data set	WSJ + Rit	90.18 %	89.44 %	89.80
	WSJ + Rit $\uparrow$	91.07 %	89.62 %	90.33
	WSJ $\downarrow$ + Rit	91.46 %	90.34 %	90.89
features	WSJ $\times$ Rit	88.98 %	88.15 %	88.56
	WSJ $\times$ Rit $\uparrow$	91.75 %	90.24 %	90.99
	WSJ $\downarrow \times$ Rit	91.49 %	90.44 %	90.96

Table: Chunking results with different models
(Tested on 408 sentences using 10-fold cross validation)

Annotation Pipeline



Manual Annotation

Annotation Procedure

Stage 1

- Identify frequent verbs (predicates)
- Group them into semantic fields
  - *cast*, *drag*, *draw*, *hang*, *lay*, *move*, *place*, *put*, *raise*, *remove*, *throw*, . . .
- Find appropriate frames
  - Placing and Moving

Annotation Procedure

Stage 1

- Identify frequent verbs (predicates)
- Group them into semantic fields
  - *cast*, *drag*, *draw*, *hang*, *lay*, *move*, *place*, *put*, *raise*, *remove*, *throw*, . . .
- Find appropriate frames
  - Placing and Moving

Stage 2

- Automatic pre-annotation of frames, based on established verb-frame lexicon
- Train automatic role labeler on manual role annotations
- Correct (unsure) role assignments manually

Annotation Process

Pre-annotation of frames

- *Offer*$_{Giving}$ *a lamp and* *speak*$_{Speaking}$ *the mantra tejo 'si.*

Annotation Process

Pre-annotation of frames

- *Offer*$_{Giving}$ *a lamp and* *speak*$_{Speaking}$ *the mantra tejo 'si.*

Manual validation of frames and role annotation

- *Offer*$_{Giving}$ *[a lamp]*$_{Theme}$ *and* *speak*$_{Speaking}$ *[the mantra tejo 'si]*$_{Text}$*.*
- Frame assignment: Between 68.8% and 91.4% precision

Annotation Process

Pre-annotation of frames

- *Offer*$_{Giving}$ *a lamp and* *speak*$_{Speaking}$ *the mantra tejo 'si.*

Manual validation of frames and role annotation

- *Offer*$_{Giving}$ *[a lamp]*$_{Theme}$ *and* *speak*$_{Speaking}$ *[the mantra tejo 'si]*$_{Text}$*.*
- Frame assignment: Between 68.8% and 91.4% precision

Semi-automatic role (and co-reference) annotation

- Role assignment: Between 50% and 100% precision depending on frequency

Overview

Aggregation and Exploitation

Abstraction and normalization of event chains

- Based on annotated frames, roles and coreference chains

Capturing statistics about "ritual elements"

- Co-occurrence of ritual objects/participants with ritual actions
- Identify overlapping (partial) event chains across rituals

Interpretation is up to ritual experts

- Omissions may disclose implicit knowledge about ritual acts
- Enable search over annotated descriptions to validate hypothesis about "ritual grammar"

Harvesting Event Chains: A Show Case

R1: *Offer a lamp (with a burning wick and the mantra) tejo 'si.*

R2: *Worship of the lamp, the wooden measuring vessel and the key (reciting) agnir mūrdhā divaḥ (and) trātāram indram. One should wave with lamp, wooden measuring vessel and key.*

R3: *Shower pieces of fruits (etc.) from the measuring vessel (on the head of the boy with) yāḥ phalini. (Make this) three (times). Show (and offer) the lamp (to the boy with) tejo 'si.*

R4: *(Wave) light (with a burning wick). Now fragrant materials etc. Worship the lamp, the wooden measuring vessel and the key (reciting) agnir mūrdhā (and) trataram indram. Wave the lamp, the (wooden) measuring vessel and the (iron) key (over the head of the boy reciting) ausraghnam.*

Harvesting Event Chains: A Show Case

R1: *Offer a lamp (with a burning wick and the mantra) tejo 'si.*

R2: *Worship of the lamp, the wooden measuring vessel and the key (reciting) agnir mūrdhā divaḥ (and) trātāram indram. One should wave with lamp, wooden measuring vessel and key.*

R3: *Shower pieces of fruits (etc.) from the measuring vessel (on the head of the boy with) yāḥ phalini. (Make this) three (times). Show (and offer) the lamp (to the boy with) tejo 'si.*

R4: *(Wave) light (with a burning wick). Now fragrant materials etc. Worship the lamp, the wooden measuring vessel and the key (reciting) agnir mūrdhā (and) trataram indram. Wave the lamp, the (wooden) measuring vessel and the (iron) key (over the head of the boy reciting) ausraghnam.*

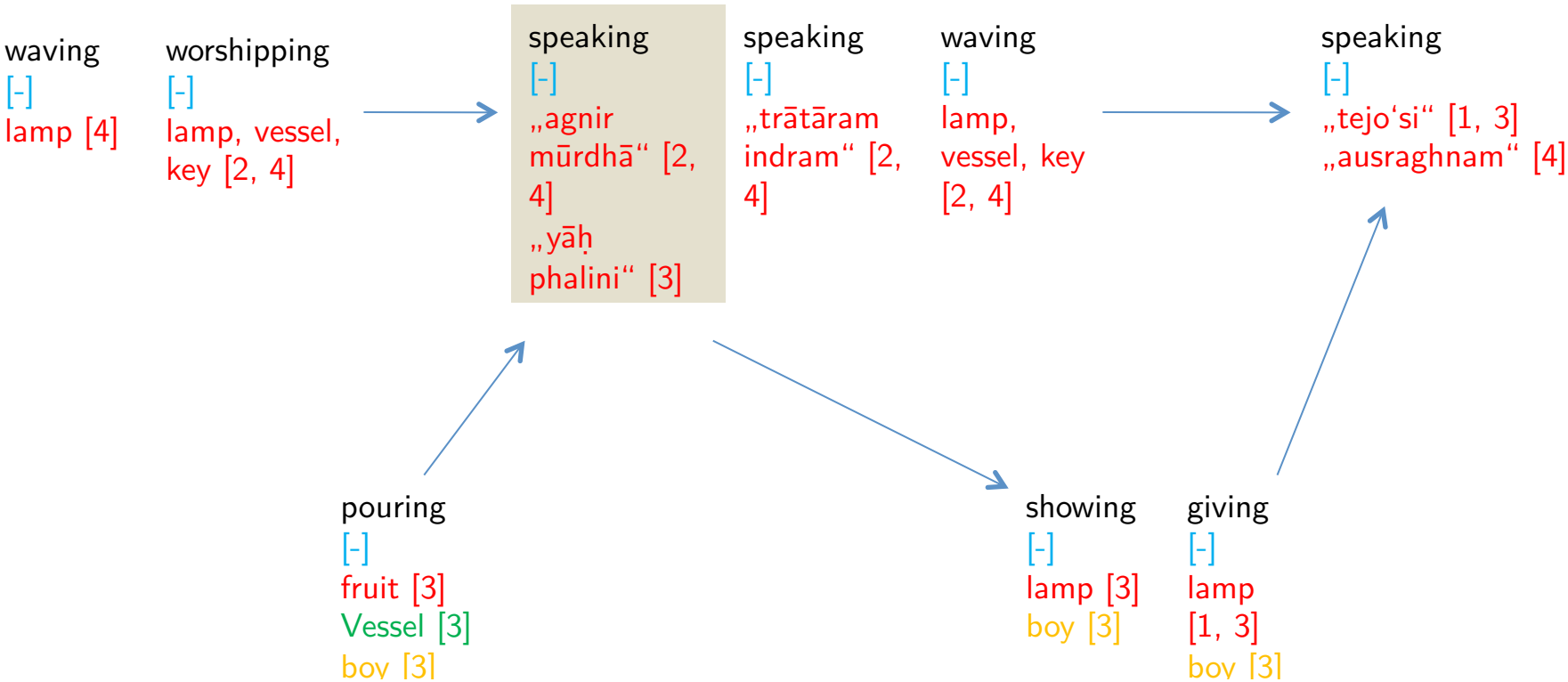
Aligning Sequences

							giving [-] lamp	speaking [-] „tejo‘si“
	worshipping [-] lamp, vessel, key		speaking [-] „agnir mūrdhā“	speaking [-] „trātāram indram“	waving [-] lamp, vessel, key			
		pouring [-] fruit Vessel boy	speaking [-] „yāḥ phalini“			showing [-] lamp boy	giving [-] lamp boy	speaking [-] „tejo‘si“
waving [-] lamp	worshipping [-] lamp, vessel, key		speaking [-] „agnir mūrdhā“	speaking [-] „trātāram indram“	waving [-] lamp, vessel, key			speaking [-] „ausra-ghnam“

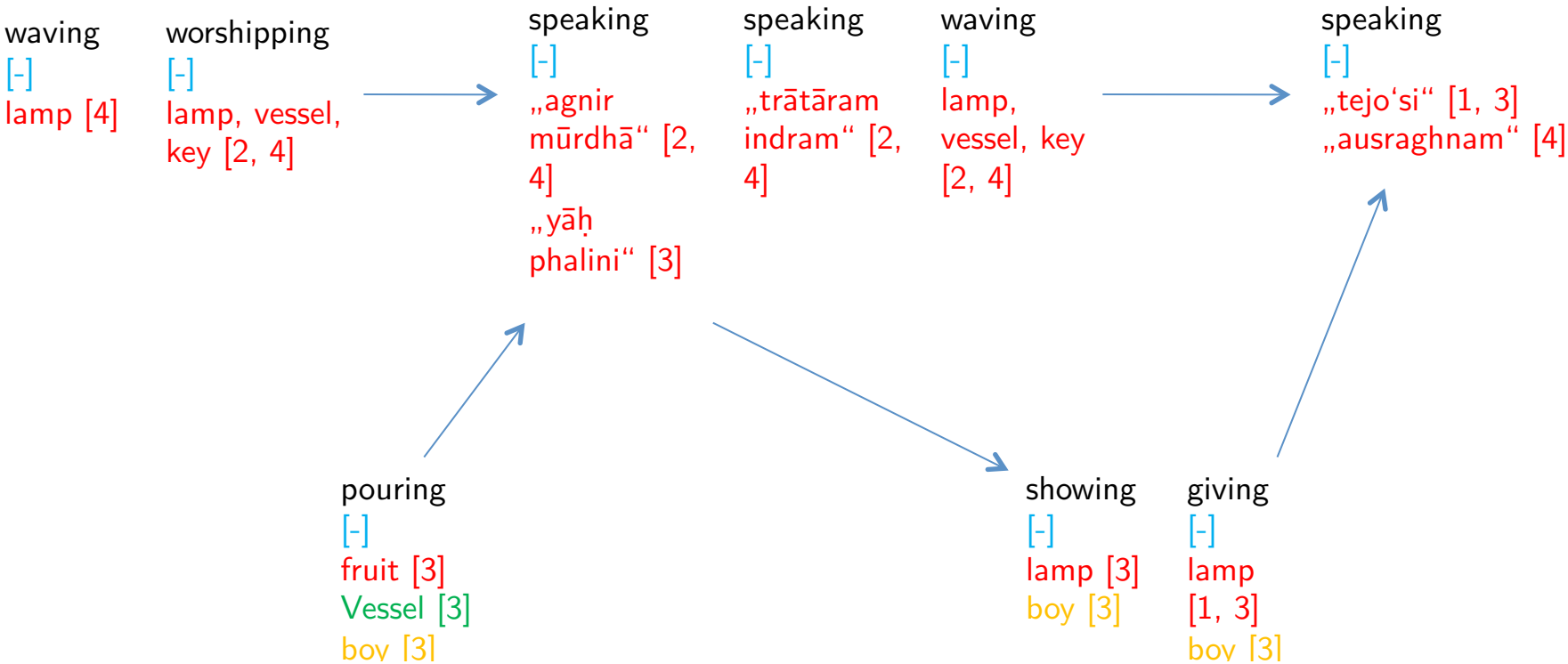
Aligning Sequences

							giving [-] lamp	speaking [-] „tejo‘si“
	worshipping [-] lamp, vessel, key		speaking [-] „agnir mūrdhā“	speaking [-] „trātāram indram“	waving [-] lamp, vessel, key			
		pouring [-] fruit Vessel boy	speaking [-] „yāḥ phalini“			showing [-] lamp boy	giving [-] lamp boy	speaking [-] „tejo‘si“
waving [-] lamp	worshipping [-] lamp, vessel, key		speaking [-] „agnir mūrdhā“	speaking [-] „trātāram indram“	waving [-] lamp, vessel, key			speaking [-] „ausra-ghnam“

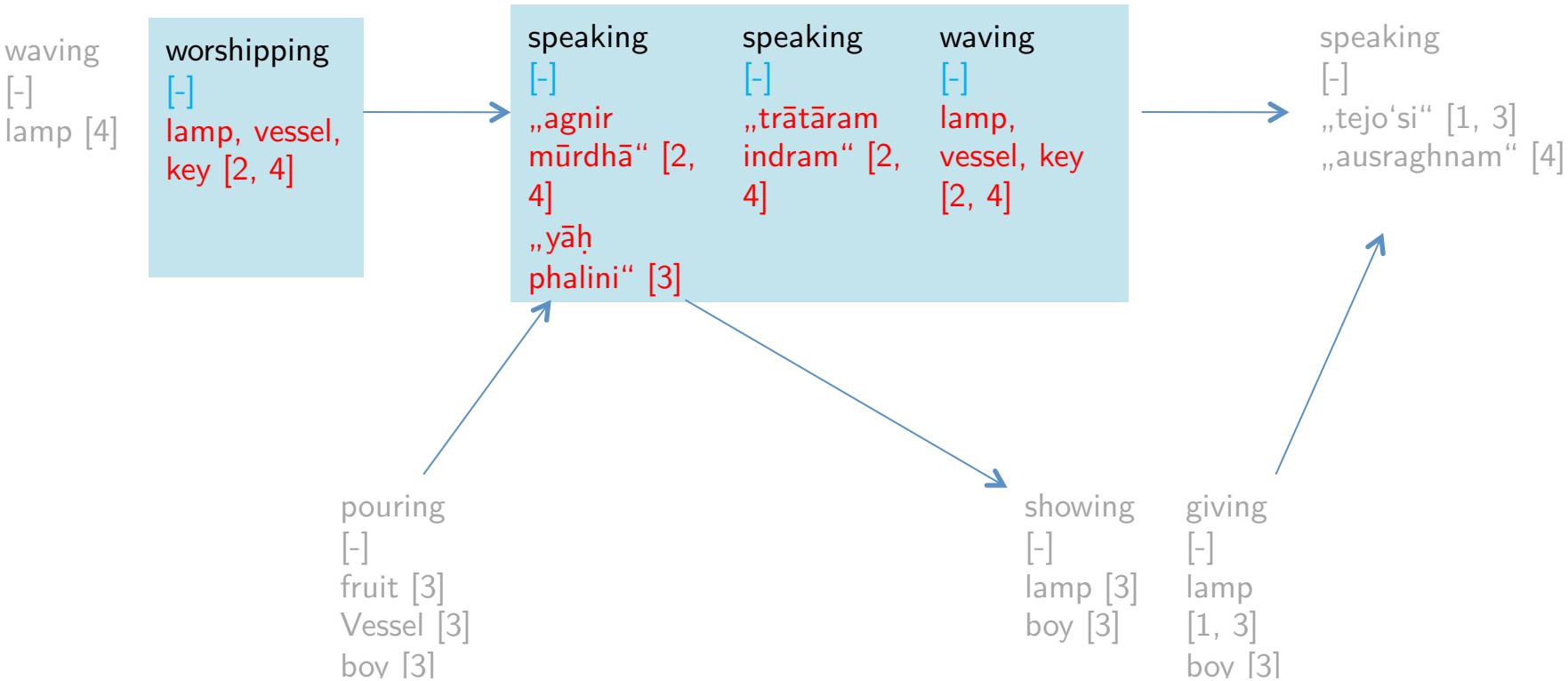
Combining Sequences



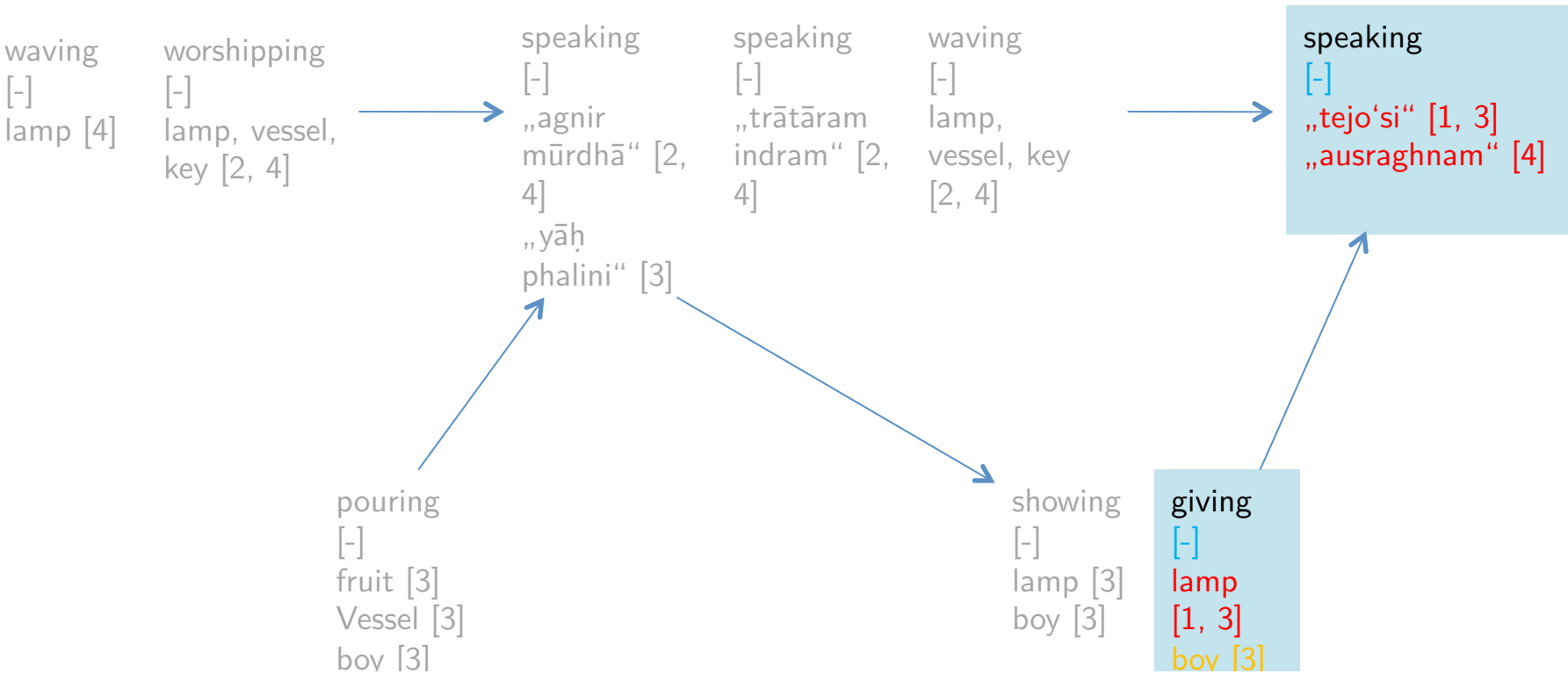
Combining Sequences



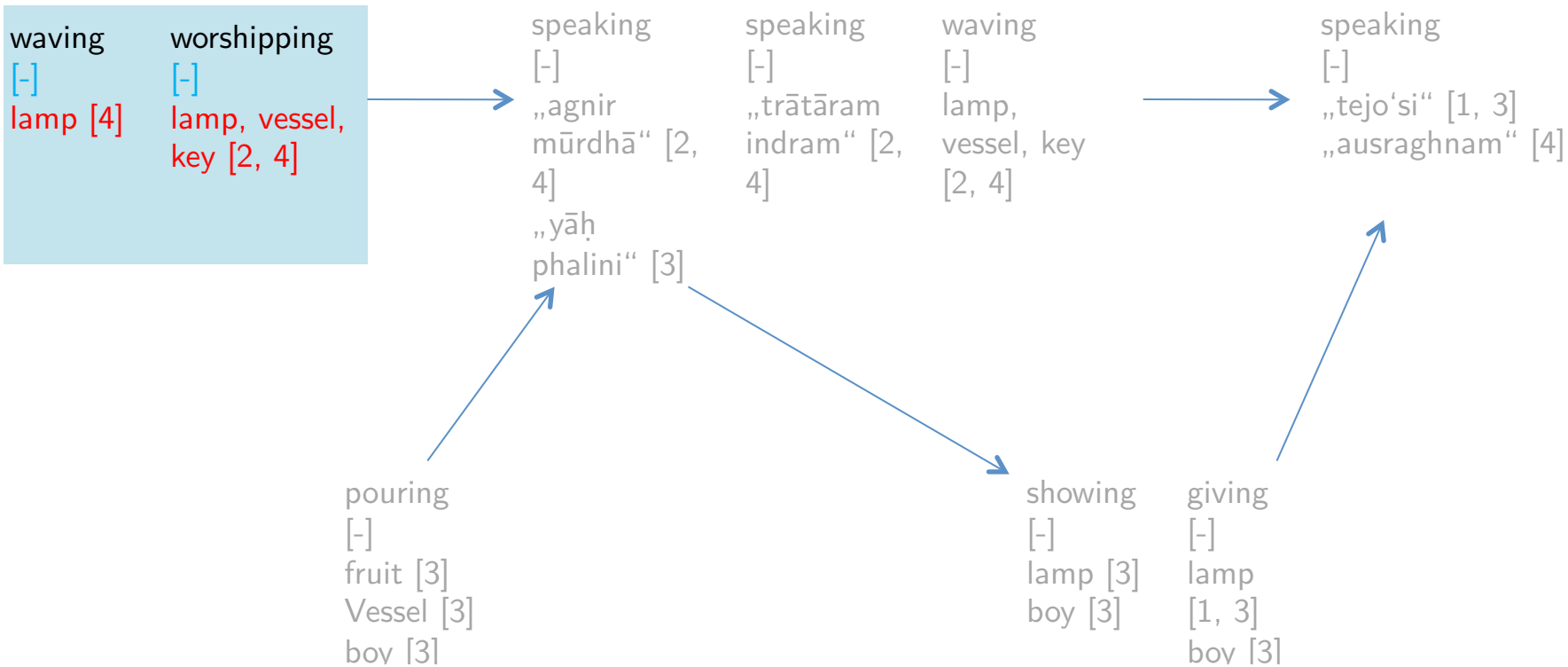
Ritual Element 1?



Ritual Element 2?



Ritual Element 3?



Calculating Bigrams

	give	speak	worship	wave	throw	pour	show
give		2				1	
speak		2		2	1		1
worship		2					
wave	1	1	1				
throw		1					
pour		1					
show	1						

Ritual Element Candidates

Co-occurring Bigrams

- give – speak
- speak – speak
- worship – speak
- speak – wave

Found by Inspection

- worship – speak – speak – wave
- give – speak

Overview

Challenges and Chances

Domain Adaptation for Preprocessing

- Harvesting events depends on *high-quality preprocessing*
- Standard tools perform poorly on the ritual domain
- *Augmenting the feature space* and *balancing training data* achieves significant performance gains

Challenges and Chances

Domain Adaptation for Preprocessing

- Harvesting events depends on *high-quality preprocessing*
- Standard tools perform poorly on the ritual domain
- *Augmenting the feature space* and *balancing training data* achieves significant performance gains

Semantic Annotation and Automation

- Frame semantics can be used to model ritual event sequences
- A large portion of FrameNet frames can be used, some frames need to be created
- *Challenges for automation*: adapting SRL, concept labeling and coreference resolution to ritual domain

Prospects

'We are not alone...'

- Domain adaptation for event-based semantic processing is crucial for many applications

Prospects

'We are not alone...'

- Domain adaptation for event-based semantic processing is crucial for many applications

Progress on all aspects of annotation

- Domain adaptation – Manual annotation – Automation

Prospects

'We are not alone...'

- Domain adaptation for event-based semantic processing is crucial for many applications

Progress on all aspects of annotation

- Domain adaptation – Manual annotation – Automation

Keep pursuing our aims

- Establish new ways of conducting empirical research in the social sciences using computational linguistic techniques

Acknowledgements

Our collaborators:

Oliver Hellwig, Anand Mishra, Axel Michaels

Our research assistants:

Irina Gossmann, Borayin Maitreya Larios, Julio Cezar Rodrigues, Britta Zeller

Thank you.

References I

Aljoscha Burchardt, Anette Frank, and Manfred Pinkal. Building Text Meaning Representations from Contextually Related Frames – A Case Study. In *Proceedings of IWCS*, 2005.

Hal Daumé III. Frustratingly easy domain adaptation. In *Proceedings of ACL*, pages 256–263, 2007. URL `http://www.aclweb.org/anthology/P07-1033`.

Hal Daumé III and Daniel Marcu. Domain Adaptation for Statistical Classifiers. *Journal of Artificial Intelligence Research*, 26:101–126, 2006.

Hal Daumé III, Tejaswini Deoskar, David McClosky, Barbara Plank, and Jörg Tiedemann, editors. *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*. Association for Computational Linguistics, Uppsala, Sweden, July 2010. URL `http://www.aclweb.org/anthology/W10-26`.

Oliver Hellwig. A chronometric approach to Indian alchemical literature. *Literary and Linguistic Computing*, 24(4):373–383, 2009.

Axel Michaels. "How do you do?" — Vorüberlegungen zu einer Grammatik der Rituale. In *Der Mensch – ein "Animal Symbolicum"? Sprache – Dialog – Ritual*. Wissenschaftliche Buchgesellschaft, Darmstadt, 2007.

References II

Axel Michaels and Niels Gutschow. *Handling Death. The Dynamics of Death Rituals and Ancestor Rituals Among the Newars of Bhaktapur, Nepal*, volume 3 of *Ethno-Indology. Heidelberg Studies in South Asian Rituals*. Harrassowitz Verlag, 2005.

Axel Michaels and Niels Gutschow. *Growing Up. Hindu and Buddhist Initiation Rituals among Newar Children in Bhaktapur, Nepal*, volume 6 of *Ethno-Indology. Heidelberg Studies in South Asian Rituals*. Harrassowitz Verlag, 2008.

Thomas Morton, Joern Kottmann, Jason Baldridge, and Gann Bierner. OpenNLP: A Java-based NLP Toolkit. http://opennlp.sourceforge.net, 2005.

Nils Reiter, Oliver Hellwig, Anand Mishra, Irina Gossmann, Borayin Maitreya Larios, Julio Cezar Rodrigues, Britta Zeller, and Anette Frank. Adapting Standard NLP Tools and Resources to the Processing of Ritual Descriptions. In Caroline Sporleder and Kalliopi Zervanou, editors, *Proceedings of the ECAI 2010 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2010)*, Lisbon, Portugal, August 2010.